Rhythm in the Kingdom: a variationist analysis of speech rhythm in Tongan English[1]

DANIELLE TOD
University of Bern

This article presents an analysis of speech rhythm in Tongan English, an emergent variety spoken in the Kingdom of Tonga. The normalised Pairwise Variability Index (nPVI-V) is used to classify the variety and determine the social and stylistic constraints on variation in a corpus of conversational and reading passage data with 48 speakers. Findings reveal a greater tendency towards stress-timing in speakers of the emergent local elite, characterised by white-collar professions and high levels of education, and those with a high index of English use. Variation is discussed as a consequence of proficiency, language contact and L1 transfer. An acoustic analysis of vowels in unstressed syllables of eight speakers confirms that lack of vowel centralisation (higher F1) is an underlying linguistic mechanism leading to more syllable-timed speech. Stark interspeaker variation was identified, highlighting the need to proceed with caution when classifying L2 Englishes based on speech rhythm.

Keywords: World Englishes, speech rhythm, variation, L2 Englishes, indexicality

1   Introduction

World Englishes are often classified on the basis of speech rhythm. Originally treated as a dichotomous typology, the classification of languages and varieties as stress-timed or syllable-timed dates back to an observation made by Lloyd James (1940), who noted that some languages exhibit *machine-gun* (syllable-timed) rhythm while others exhibit *Morse-code* (stress-timed) rhythm. This distinction was elaborated on by early phoneticians including Pike (1945) and Abercrombie (1967), who defined stress-timing and syllable-timing as follows: *stress-timed* varieties are those in which syllable duration is varied but stress appears at roughly equal intervals, while in *syllable-timed* varieties, syllables occur at equal intervals and are roughly the same in duration. Today, stress-timing and syllable-timing are considered tendencies rather than a strict dichotomy and World Englishes are classified with reference to this continuum (e.g. Mesthrie 2008; Mesthrie & Bhatt 2008; Fuchs 2016a). Since the turn of the century, various acoustic metrics have been developed by scholars in the field of sociophonetics for the quantification of speech rhythm that provide scores indicating degree of stress-timing versus syllable-timing (see Fuchs 2016a: 39–41, for an

extensive list). In the current study, I employ one of these metrics, the normalised Pairwise Variability Index (nPVI-V), to examine variation in speech rhythm in a corpus of recorded conversational and read speech with 48 speakers of Tongan English (hereafter TE), an emergent variety spoken in the Kingdom of Tonga.

Tonga is an independent nation state and constitutional monarchy in Polynesia comprising an archipelago of over 150 islands roughly 2,000km north of New Zealand and 3,300 km east of Australia in the South Pacific (Riches & Stalker 2016). The population is rather homogeneous, with 97 per cent of residents identifying as Tongan and the remaining 3 per cent as part Tongan, Samoan, Fijian or other (Tonga Statistics Department 2019). A traditional hierarchy is maintained that distinguishes between royalty, nobility and commoners, with a fourth tier of local 'elite' commoners having emerged in recent decades characterised by an advantageous social or economic position (Moala 2009). This group includes the likes of the educated, clergy, *pālangi* (Caucasian, often Anglophone, foreigners) and those who are successful in commerce or in a leadership position (Bott 1981; Moala 2009).

Official language status has been granted to both Tongan and English (Taumoefolau 1998), with Tongan the L1 of the vast majority used for daily interactions while English is mostly acquired as an L2 through formal education. English is beginning to encroach on the private sphere; it is used more frequently in the home as it is viewed as the gateway to opportunities and upward social mobility. In particular, local elite use English in an increasing number of domains, providing a linguistic model for upwardly mobile Tongans (Taumoefolau 1998) and reinforcing the status of English as a prestige language. Local elite, including royalty and nobility, are often more proficient English speakers as they have the financial means to enrol their children in English-medium schools, both locally and abroad. Therefore, a great deal of variation exists across TE speakers regarding degree of contact with and daily use of English, leading to varying levels of English language proficiency and differences in speech rhythm.

The current study aims to classify TE on the basis of speech rhythm, placing the variety within the World Englishes landscape; it examines the social and stylistic constraints on variation in TE, providing a nuanced picture of inter- and intraspeaker variation in speech rhythm. In doing so, I highlight the potential second-order indexicality (Silversteen 2003) and social meaning of speech rhythm in the local Tongan context. The acoustic nPVI-V metric is demonstrated as a fruitful tool for observing gradient differences in speech rhythm both within and between speakers, provided that a principled and consistent approach to segmentation is maintained. Application of nPVI-V provides a nuanced understanding of speech rhythm for the classification of World Englishes and local social meaning tied to rhythmic variation in contrast to impressionistic accounts. Finally, the current article demonstrates a relationship between syllable-timing and vowel reduction in unstressed syllables as an underlying linguistic mechanism.

1.1 Speech rhythm in World Englishes

Research on speech rhythm in World Englishes reveals an overwhelming tendency for inner-circle (L1) varieties to be classified as stress-timed while emergent and L2 varieties are often deemed syllable-timed. This classification is often based on impressionistic or auditory analyses, although studies over the past two decades have begun to include quantitative and acoustic approaches. Varieties termed *stress-timed* include Standard Southern British English (SSBE), as well as British English dialects such as Bristolian, Welsh Valleys, Orkney Islands and Shetland Englishes (Knight 2011; White & Mattys 2007a). Although variation between British dialects has been found, on the whole, these varieties remain on the stress-timed end of the continuum. Moving beyond the British Isles, Thomas & Carter (2003, 2006) found that General American (GenAm) and African American English (AAE) are equally stress-timed, while ex-slave AAE is more 'intermediate' on the rhythm continuum. In New Zealand, Holmes (1997) found that Pākeha (New Zealand European) English is more stress-timed than Māori English, which is attributed to a transfer effect from syllable-timed Māori. Szackay (2006) also examined rhythm in Pākeha and Māori Englishes using acoustic metrics, finding quantitative evidence for more syllable-timed tendencies in the latter while speakers of the former exhibited stress-timed tendencies close to speakers of Standard British English (BrE).

Emergent and L2 Englishes deemed more *syllable-timed* include East African, Nigerian, Ghanaian, Indian, Pakistani and Philippine Englishes (Mesthrie & Bhatt 2008: 129), Singapore English (Low & Grabe 1995, 1999; Low, Grabe & Francis 2000; Deterding 2001), Brunei English (Deterding 2012), Taiwanese English (Jian 2004a, 2004b) and Korean English (Kim, Flynn & Oh 2006). Emergent and L2 varieties of English in the Pacific are also overwhelmingly syllable-timed. In Micronesia, syllable-timed varieties include Palauan English (Britain & Matsumoto 2012) and Saipan English (Hess 2019); in Guam English, syllable-timing is characteristic of older and basilectal speakers while younger speakers tend to be more stress-timed (Kuske 2019). Further syllable-timed varieties spoken in Polynesia and Melanesia include Fiji English (Tent & Muggler 2008), Samoan English (Biewer 2020) and Māori English (Ainsworth 1993; Bauer 1994; Holmes 1997; Szackay 2006).

A tendency towards stress-timing has been associated with vowel reduction and centralisation in unstressed syllables (e.g. Mesthrie & Bhatt 2008; Deterding 2012; Kuske 2019), with centralisation shown to result in shorter duration (Lindblom 1963; Beinum 1980; Kim & Lee 2005), leading to greater variation in duration and the perception of stress-timing. Empirical evidence supporting this observation has been provided by Jian (2004b), who examined the formant space in speakers of GenAm (L1) and Taiwanese (L2) English. Reading artificial sentences in which some included only full vowels and others included reduced vowels, acoustic analysis revealed that GenAm speakers made more use of the central vowel space than Taiwanese English speakers. In a parallel study, speech rhythm was examined acoustically in the same speakers, whereby Taiwanese English speakers exhibited more syllable-timed

tendencies, providing evidence for a relationship between vowel centralisation in unstressed syllables and speech rhythm (Jian 2004a). Lack of centralisation and reduction in L2 English are often cited as a transfer effect, whereby speakers exhibit transfer from an L1 phoneme inventory that does not contain fully reduced vowels. Further support for this argument comes from a study on Thai English (Sarmah, Gogoi & Wiltshire 2009), which was found to exhibit stress-timed rhythm similar to BrE using quantitative acoustic metrics, despite being an L2 variety. The authors note that this is the result of an L1 transfer effect from Thai, which is stress-timed and exhibits contrast in vowel length and quality. Finally, Nishihara & Van de Wiejer (2011) argue that syllable-timing is unmarked and simpler for L2 speakers to acquire, contributing to the prevalence of syllable-timing in L2 varieties.

Several studies on speech rhythm in varieties of English and other languages have focused specifically on the degree to which L2 speakers of English acquire native-like patterns in rhythm, and the extent to which this is influenced by rhythm in the L1. Findings consistently show that speakers of a syllable-timed L1 exhibit intermediate rhythmic patterns in L2 English (e.g. Lee & Kim 2005; Kim *et al.* 2006; White & Mattys 2007a, 2007b), pointing to a transfer effect on L2 rhythm acquisition. For example, Carter (2005) examined rhythm in bilingual Spanish and English speakers using a quantitative acoustic metric. Spanish speakers of English as an L2 were found to exhibit intermediate metric scores between those of L1 English and L1 Spanish speakers, providing evidence for L1 transfer effects on rhythm in L2 speakers and new dialect formation. This transfer phenomenon has also been identified in L1 speakers of a stress-timed variety who acquire a syllable-timed L2. Among other prosodic properties, Bond, Markus & Stockmal (2003) applied quantitative metrics to a corpus of read speech data from native speakers of Latvian (syllable-timed) and learners of Latvian with Russian (stress-timed) as an L1. Metric scores were similar for proficient learners and native speakers of Latvian, while less proficient speakers had higher metric scores, evidencing stress-timed L2 speech and thus a transfer effect from the L1, results which have been replicated by Stockmal, Markus & Bond (2005). In a study by White & Mattys (2007a, 2007b), no significant rhythmic difference was found in L1 speakers of (stress-timed) Dutch when using their native tongue in relation to rhythm in L2 English, and vice versa, indicating that rhythmic differences in L2 speech are at least in part a function of L1 rhythm. On the other hand, L2 speakers of English with (syllable-timed) Spanish as an L1 exhibited a tendency towards syllable-timing. Importantly, this indicates that syllable-timing in emergent and L2 Englishes can be attributed at least in part to L1 transfer as opposed to second language acquisition processes. Given that TE is an emergent L2 variety, the rhythmic properties of Tongan as an L1 must be taken into account and are explored below.

1.2 *Rhythm and prosody in Tongan*

Speech rhythm in the Tongan language has not received any attention in the literature. Syllable structure, stress and vowel quality, on the other hand, have been more

thoroughly described, and provide clues as to what we may expect to see in TE, given the L1 transfer effect proposed in the literature outlined above. Tongan has a rigid syllable structure, allowing only open syllables of CV and V. While early phonological accounts of Tongan argued that a distinction can be made between long and short vowels, i.e. V and V′ (e.g. Biggs 1971; Feldman 1978), more recent research suggests that so-called 'long vowels' or 'diphthongs' are best described as sequences of two or more syllables containing one vowel each (e.g. Taumoefolau 1998; Anderson & Otsuka 2006). The Tongan vowel inventory includes five phonemes, namely /a, e, i, o, u/ (Feldman 1978: 133), and no central vowels.

Tongan carries primary stress on the penultimate syllable of a stress group (Garellek & Tabian 2019: 7). Stress is therefore a property of an environment whereby the syllable is the stress-bearing unit, and not the grammatical or semantic word (Taumoefolau 1998). The 'end rule right' means that additional prominence is given to the rightmost stressed syllable in a phonological phrase (Anderson & Otsuka 2006). Primary stress is evident in vowels by measures of F0, intensity, duration, F1 and voice quality, the strongest of which are F0 and duration (Garellek & White 2012, 2015). In an examination of vowel space in Tongan speakers, Garallek & White (2015) note that centralisation is not characteristic of vowels in unstressed syllables despite lower F1 values in all five vowels, as vowel space does not significantly change.

The absence of vowel centralisation in unstressed syllables in Tongan and the link between stress and environment as opposed to word class would suggest that Tongan exhibits syllable-timed tendencies, while the significance of duration as a marker of stress suggests that speakers are sensitive to vowel length. Given the evidence pointing to L1 transfer effects in L2 rhythm, I predict that we will see a general tendency towards syllable-timing in TE in relation to inner-circle standard varieties.

1.3 Quantifying speech rhythm

Since the turn of the century, numerous acoustic metrics for quantifying speech rhythm have been developed by scholars in the field of sociophonetics, of which a favoured and oft-applied metric is the Pairwise Variability Index (PVI), devised by Low & Grabe (1995, 1999) and expanded by Low et al. (2000). In the current study, I employ an elaboration of the PVI, namely the speech-rate normalised **nPVI-V**, devised by Low (1998, as cited in Low 2006), the formula for which is:[2]

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k - d_{k-1})/2} \right| / (m-1) \right]$$

PVI metrics calculate the mean difference in duration of intervals between successive interval pairs, based on the hypothesis that there is less variability in duration of

---

[2] m = the number of vocalic intervals and d_k = the duration of the kth vocalic interval.

successive vocalic and consonantal intervals in syllable-timed than in stress-timed varieties. The nPVI-V is a speech-rate normalised version of the PVI metric in which successive vocalic (vowel) interval durations are measured, first used by Low (1998, as cited in Low 2006), who applied the metric to speakers of Singapore English (more syllable-timed) and BrE (more stress-timed). The output of the nPVI-V metric is an index score, which can be used to place varieties in terms of a rhythmic continuum: *higher* scores indicate greater variability and thus greater tendency towards *stress-timing*, while *lower* scores are indicative of *syllable-timing*. Low found a significant difference in scores between the two varieties, providing support for the aforementioned classification.

Since its development, the nPVI-V has been applied to a number of L1 and L2 varieties, as illustrated in table 1. Overall, higher scores are evident across L1 standard varieties, exhibiting a tendency towards stress-timing, while lower scores are evident across many emergent and L2 Englishes, exhibiting a tendency towards syllable-timing. It is also important to bear in mind whether the nPVI-V scores were applied to spontaneous or read speech, as different speaking conditions has been found to predict intraspeaker variation (e.g. Roach 1982; Arvaniti 2012).

The nPVI-V metric belongs to an extensive range of metrics developed over the past two decades for the quantification of speech rhythm (see Fuchs 2016a: 39–41 for an extensive list), but the nPVI-V has been the most widely applied and attested as

Table 1. *Mean nPVI-V scores for several World Englishes in both read and spontaneous speech*

Variety	Speech style	Mean nPVI-V	Reference
Brunei English	Read	40.5	(Deterding 2012: 10)
Hispanic English	Spontaneous	42.7	(Carter 2005: 69)
Malaysian English	Spontaneous	43.7	(Tan & Low 2014: 209)
Māori (New Zealand) English	Spontaneous	47.3	(Szackay 2006: 425)
Singapore English	Spontaneous	48.1	(Tan & Low 2014: 209)
London (non-Anglo Hackney) English	Spontaneous	51.1	(Torgensen & Szakay 2011: 169)
Indian English	Spontaneous	52.4	(Fuchs 2016a: 114)
European American English	Spontaneous	53.0	(Carter 2005: 69)
London (Anglo Hackney) English	Spontaneous	53.3	(Torgensen & Szakay 2011: 169)
African American English	Spontaneous	55.2	(Carter 2005: 69)
Indian English	Read	55.6	(Fuchs 2016a: 114)
Thai English	Spontaneous	56.5	(Sarmah *et al.* 2009: 206)
British English (unspecified variety)	Spontaneous	58.3	(Fuchs 2016a: 114)
Pākehā (New Zealand) English	Spontaneous	58.7	(Szackay 2006: 425)
British English	Read	61.3	(Fuchs 2016a: 114)

somewhat more 'robust' across the literature. Low *et al.* (2000) compared PVI metrics to further metrics such as the standard deviation measures ΔV and ΔC proposed by Ramus, Nespor & Mehler (1999), whereby scores were calculated for 100 utterances each of read speech by BrE and Singapore English speakers. Based on their findings, they argue that the normalised nPVI is the most reliable metric of rhythmicity, as it is cumulative and controls for changes in speech rate. In a study on rhythm in stress-timed languages (English and Dutch) and syllable-timed languages (Spanish and French), White & Mattys (2007a) found nPVI-V (in addition to %V and varcoV) to most accurately distinguish these languages. The same metrics were also informative as to the adaptation of speakers to rhythmically similar (Dutch and English) or rhythmically distinct (Spanish and English) varieties. In a summary of findings from previous research on rhythm, Fuchs (2016a: 56) also notes that nPVI-V is a preferable measure as it is 'robust to variation in speech rate and relatively robust to variation in sentence materials, speakers and transcribers'. Non-rate-normalised metrics are less reliable as they do not discriminate as well between languages and speakers. It is for these reasons that the nPVI-V was chosen as an acoustic metric for the current study. The application of the metric and data used for analysis are presented in the following section.

2 Data and method of analysis

2.1 Participants

Speech rhythm in TE was analysed in a conversational condition and a reading passage condition, for which the composition of the corpus differs slightly. A corpus of informal conversations with 48 speakers was used for the analysis of rhythm in the conversational condition, recorded in 2019 on a three-month fieldwork trip to Tongatapu, the main island of the Tongan archipelago. Speakers in this corpus include 25 women and 23 men born between 1929 and 2003 (aged 16–90, mean age = 45.3) who identify as Tongan and represent a cross-section of the local English-speaking population in Tonga. This means that the corpus includes a disproportionate number of more educated speakers in relation to the population average, with 26 speakers currently undertaking or having completed tertiary education, 12 currently undertaking or having completed secondary school as their highest level of education and 10 for whom primary school is their highest level of education. These speakers were selected from a larger corpus of recorded TE speech to reflect a balance in terms of locally salient external constraints including *age*, *sex*, *education*, *occupation* and *index of English use*, outlined in section 2.2 below. Speaker selection was further limited to those individuals who had produced enough fluid speech in the conversational condition, devoid of hesitations and pauses, such that a sufficient number of utterances could be extracted for analysis.

In order to examine speech rhythm in the reading passage condition, a subcorpus containing 43 of these 48 speakers was used, in which participants read aloud a short reading passage, namely an adapted version of Aesop's fable 'The North Wind and the

Table 2. *External constraints considered in the analysis of rhythm in Tongan English*

Factor	Factor type	Factor levels	Number of participants
Age	Continuous	Ages ranging from 17–90	All
Sex	Categorical	Men	23/48
		Women	25/48
Occupation	Categorical	Student	6/48
		Blue-collar	15/48
		White-collar	27/48
Education	Categorical	Primary	10/48
		Secondary	12/48
		Tertiary	26/48
Index of English use	Continuous	Scores ranging from 4–20	All

Sun'.[3] This subcorpus includes 26 women and 19 men (aged 16–90, mean age = 44.7). The need for a subcorpus arose as not all of the 48 speakers in the conversational condition produced sufficiently fluid speech in the reading passage condition such that rhythm metrics could be applied.

2.2   External constraints

Five potential external constraints on variation in rhythm in TE were identified and included in analysis, namely *age*, *sex*, *occupation*, *education* and *index of English use*, illustrated in table 2, including the type of variable, factor levels and the number of speakers for each factor level. *Age* is treated as a continuous variable, while *sex* is treated as a binary categorical variable. The constraints *level of education* and *occupation type* serve as proxies for social background, where the distinction between blue-collar and white-collar occupation is based on primary position of employment reported by speakers in the interview. Recent decades have seen the emergence of an elite group of commoners in Tonga characterised by an advantageous socioeconomic position, namely white-collar positions and high levels of education. These proxies therefore capture a distinction between local elite and other commoners.

Turning to *index of English use*, this quantitative index was devised to capture the degree to which a speaker actively uses (or has used) English in daily communication. Speakers received an index score ranging from 4 to 20, which was calculated on the basis of English use across four key domains identified as relevant in the Tongan context, namely work, education, the home and abroad. Although media constitutes a fifth domain of English use, this was not included in the index calculation as the vast majority of speakers in the corpus indicated a preference for English in this domain. It

---

[3]   While this passage is widely used for illustrations of the IPA, several phonemes are absent or occur only in restricted settings such as /ʒ/, word-initial and medial /z/, word-initial /θ/ and dark /l/. I thus added a sentence to the passage that includes these phonemes.

is important to bear in mind that a high index score does not necessarily imply contact with native varieties of English, as speakers may be using English in interaction with other TE-speaking interlocutors.

2.3 Application of nPVI-V

In order to apply nPVI-V, a minimum of ten utterances were extracted from the conversational speech data, each containing a minimum of eight syllables. In several cases, more than ten utterances were selected in order to ensure that a minimum of 200 intervals could be analysed per speaker. A minimum of five utterances containing at least eight syllables each were also extracted from the reading passage data in order to examine the effect of attention to speech on rhythm, resulting in a minimum of ninety (vocalic and consonantal) intervals per speaker. Utterances were chosen in which speech was fluid, uninterrupted and devoid of hesitations or pauses. Following Thomas & Carter (2003: 3), a pause was considered a period of silence lasting ≥100ms. No utterances were extracted from the first five minutes of interview speech, in order to minimise the effect on speech rhythm of discomfort or nervousness due to the presence of a recording device. Further exclusions included utterances in which code switching occurred or non-lexical material such as 'um' or 'eh'. Utterances containing approximants (/w, j, ɹ, l/) were included despite the difficulty this poses for segmentation, as portions of free speech devoid of approximants are very rare. Following White & Mattys (2007b: 507), pre-pausal and utterance-final intervals were included in analysis, despite the possibility of lengthening effects, as lengthening is also common within utterances and is worthy of inclusion as it may contribute to perceptions of rhythmicity across speakers and varieties. Nevertheless, upper and lower limits for interval durations were enforced such that each interval had a maximum duration of 400ms and a minimum of 25ms. Deterding (2012) recommends imposing a minimum limit on interval duration in order to account for vowel omission in connected speech. Indeed, some reduced vowels were somewhat 'swallowed' in the data, with formants not apparent in the spectrograms despite being evident auditorily.

Before quantitative metrics could be applied, utterances were segmented into vocalic and consonantal intervals in Praat (Boersma 2001). The forced-alignment tool 'WebMAUS' (Kisler, Reichel & Schiel 2017) was used to automatically segment each utterance according to word and phone boundaries, the output of which (in .TextGrid format) was used to manually place interval boundaries in accordance with segmentation criteria proposed by Peterson & Lehiste (1960: 694–8) and successive vocalic and intervocalic intervals were combined. The placement of boundaries between vocalic and intervocalic intervals is a thorny process as the border between intervals is often unclear in the spectrogram, particularly in the case of approximants (/w, j, ɹ, l/). In order to ensure consistency, boundary placement was done by the author alone. Following Thomas & Carter (2003, 2006) and Ramus et al. (1999), I treated approximants as consonantal in syllable onset position and vocalic in syllable coda position, unless there was a clear change in formant structure visually evident in the

spectrogram. For example, in 'north wind', /w/ was treated as a consonantal interval, while in 'stronger than', /ɹ/ was treated as vocalic unless there was a clear change in formant structure. Furthermore, glottalised intervals were treated as consonantal intervals. In order to determine which approximants belonged to the syllable onset or coda, I followed the maximum onset principle, which asserts that consonants belong to the following syllable so long as this does not violate the phonotactic constraints of syllable onsets according to language-specific conditions.

Finally, metrics were applied using a Praat script titled 'durationAnalyzer. praat', available under www.pholab.uzh.ch/static/volker/software/plugin_durationAnalyzer. zip. This script runs through the corpus utterance-by-utterance, producing a table with a resulting value for each metric per utterance. The output of this script is values for each utterance from which mean metric scores for each speaker were calculated.

3   Results

3.1   Rhythm in conversational Tongan English

The mean nPVI-V score for conversational TE across the corpus (N = 48) is 47.9, as illustrated in table 3 in relation to other varieties of English for which mean nPVI-V scores in spontaneous or conversational speech are known. TE exhibits a lower mean nPVI-V score than most native varieties including GenAm (reported in table 3 as European American English), AAE, BrE, London (Anglo Hackney) English and Pākehā NZE, as well as several non-native varieties such as Singapore English and Thai English, indicating a more syllable-timed tendency. The mean nPVI-V score for TE is higher than Hispanic English, Malaysian English and Māori (New Zealand)

Table 3. *Mean nPVI-V scores for conversational Tongan English across the corpus (N = 48) in relation to other World Englishes*

Variety	Mean nPVI-V	Reference
Hispanic English	42.7	(Carter 2005: 69)
Malaysian English	43.7	(Tan & Low 2014: 209)
Māori (New Zealand) English	47.3	(Szackay 2006: 425)
Tongan English	**47.9**	
Singapore English	48.1	(Tan & Low 2014: 209)
London (non-Anglo Hackney) English	51.1	(Torgensen & Szakay 2011: 169)
Indian English	52.4	(Fuchs 2016a: 114)
European American English	53.0	(Carter 2005: 69)
London (Anglo Hackney) English	53.3	(Torgensen & Szakay 2011: 169)
African American English	55.2	(Carter 2005: 69)
Thai English	56.5	(Sarmah *et al.* 2009: 206)
British English	58.3	(Fuchs 2016a: 114)
Pākehā (New Zealand) English	58.7	(Szackay 2006: 425)

English, and closest to Singapore English (48.1) and Māori English (47.3). TE thus sits towards the syllable-timed end of the continuum, mirroring the overwhelming tendency for emergent and L2 varieties to exhibit more syllable-timed tendencies than inner-circle L1 varieties.

Mean nPVI-V scores across individual speakers of TE reveal a great deal of interspeaker variation, with scores ranging from 35.7 (rather syllable-timed) to 61.3 (rather stress-timed). Some speakers exhibit nPVI-V scores and rhythmicity closer to each pole of the continuum as opposed to the mean nPVI-V score for TE. This finding is illustrated in figure 1, which plots mean nPVI-V scores for individual speakers in relation to the mean nPVI-V score across the full sample (N = 48) of conversational TE. This supports the conceptualisation of rhythm as a continuum as opposed to a dichotomy, a favoured perspective in recent decades of speech rhythm research. Importantly, speakers of TE exhibit a great deal of variability in rhythm and can be placed at various points on the continuum. The significance of this result and the external constraints on this variation are explored in the following section.

3.2   External constraints

In order to examine the effect of external constraints on speech rhythm in TE, a mixed-effects linear regression model was generated for the conversational data in R using the package 'lme4' (Bates *et al.* 2015). The first model included all five external
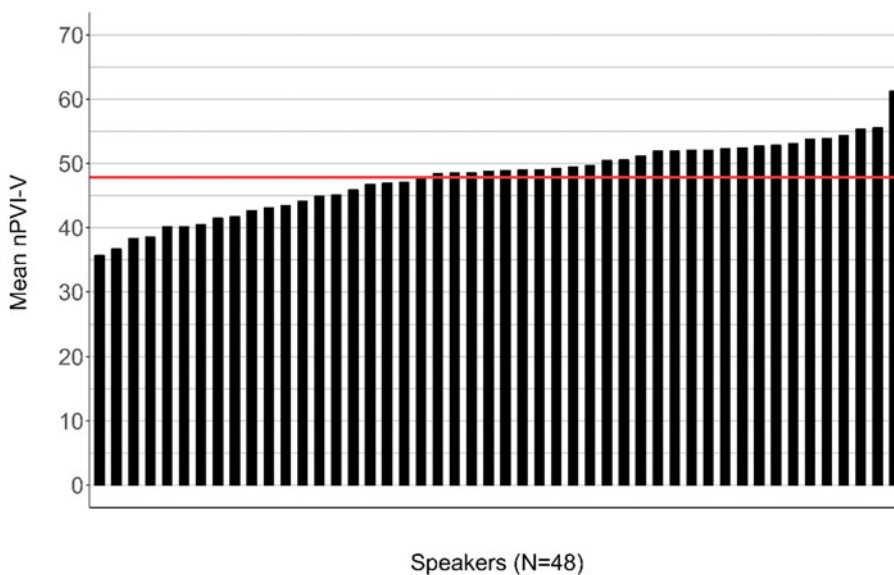

Figure 1. Mean nPVI-V scores for individual speakers of Tongan English (N = 48) in informal conversational speech. Higher nPVI-V scores indicate tendency towards stress-timing, while lower nPVI-V scores indicate tendency towards syllable-timing. The horizontal line = mean nPVI-V for whole sample.

constraints and *speaker* as a random intercept. The presence of collinearity between factors was assessed by applying the Variance Inflation Factor (VIF) and, as expected, *education* and *occupation* as proxies of a speaker's social background were highly collinear (*education* VIF = 13.57, *occupation* VIF = 12.13). This can be explained by the fact that the majority of blue-collar workers reported primary school as their highest level of education, while the majority of informants in a white-collar profession had a tertiary level education. Subsequent linear regression models thus included only *occupation*, bearing in mind that this is collinear with *education*. Models were manually stepped down with goodness of fit assessed via the application of likelihood ratio tests (anova) and comparison of AIC (Aikake Information Criterion) values. The best-fit model is illustrated in table 4, containing *occupation* and *index of English use* as significant external constraints on nPVI-V ($p < .001$).

Regarding *occupation*, students (mean nPVI-V = 49.9) and white-collar professionals (mean nPVI-V = 50.3) are more likely to produce higher nPVI-V scores than those in a blue-collar profession (mean nPVI-V = 42.8). A pairwise comparison was calculated for *occupation* using the package 'emmeans' (Lenth 2018), which revealed no significant difference between mean nPVI-V values for students and white-collar workers. Figure 2 exhibits median and range of nPVI-V scores for occupation. Essentially, these results indicate that blue-collar workers exhibit more syllable-timed tendencies in conversational speech while students and white-collar workers exhibit more stress-timed tendencies in this condition. With regards to *education*, those with a tertiary level of education are more likely to exhibit more stress-timed speech, while those with a primary or secondary level of education are more likely to exhibit syllable-timed speech. It should be noted that of the six speakers who comprise the student group, four were currently enrolled in tertiary level programmes while the remaining two were aspiring tertiary students in their final years of secondary school.

Table 4. *Best-fit linear mixed effects regression model for external constraints on nPVI-V in conversational Tongan English*

Number of observations = 489, Random effect of speaker (SD = 1.707)

Factor group	Levels	Est.	SE	df	t-value	p-value		Mean nPVI-V
(Intercept)		34.777	2.371	46.698	14.667	< 2e-16	***	–
Occupation	*(ref. level: Blue-collar)*							42.8
	Student	5.459	1.843	46.056	2.962	0.004820	**	49.9
	White-collar	4.151	1.496	47.981	2.775	0.007853	**	50.3
Index	–	0.878	0.237	48.014	3.704	0.000547	***	–

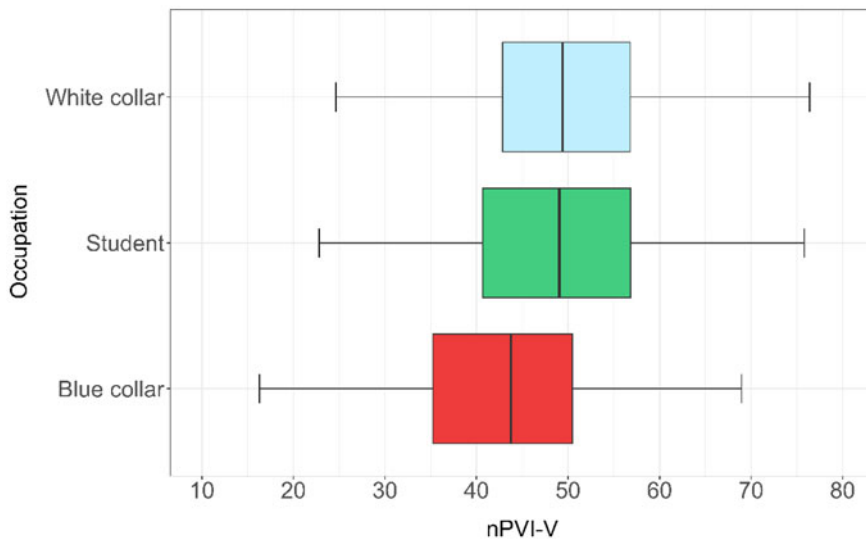Significance codes: * < 0.05, ** < 0.01, *** < 0.001

Figure 2. Median and range of nPVI-V scores for speakers of Tongan English (N = 48) in informal conversational speech according to occupation. Higher nPVI-V scores indicate tendency towards stress-timing, while lower nPVI-V scores indicate tendency towards syllable-timing.

The metric of nPVI-V is thus sensitive to social background, with the educated and white-collar 'elite' more towards the stress-timed end of the rhythm continuum.

Results for *index of English use* indicate a positive linear relationship whereby nPVI-V scores increase as index score increases. This relationship is illustrated in figure 3, which plots nPVI-V scores for each observed utterance. It is evident that the higher the index score of a speaker, the more stress-timed their speech is likely to be. In other words, the more a speaker actively uses (or has used) English in domains including the home, work, abroad and education, the higher the variability in duration of successive vocalic intervals and the more stress-timed speech rhythm will be. The mean nPVI-V values also support this finding, as speakers with the lowest index score (6) exhibit a mean nPVI-V of 39.1 and speakers with the highest index score (18) exhibit a mean nPVI-V of 61.3.

3.4 Conversational versus read speech conditions

In order to explore the effect of speech condition on rhythm, a linear regression model was generated for nPVI-V as an outcome variable in which *speech condition* was included as a fixed effect and *speaker* conditioned on *speech condition* as a random intercept. *Speech condition* did not return as a significant constraint on nPVI-V ($p = .143$) therefore speech rhythm in TE is not affected by degree of attention paid to speech. Mean nPVI-V scores for individual speakers were nevertheless plotted in order to observe possible intraspeaker variation, as seen in figure 4, illustrating that mean nPVI-V scores were higher in the reading condition than the conversational condition
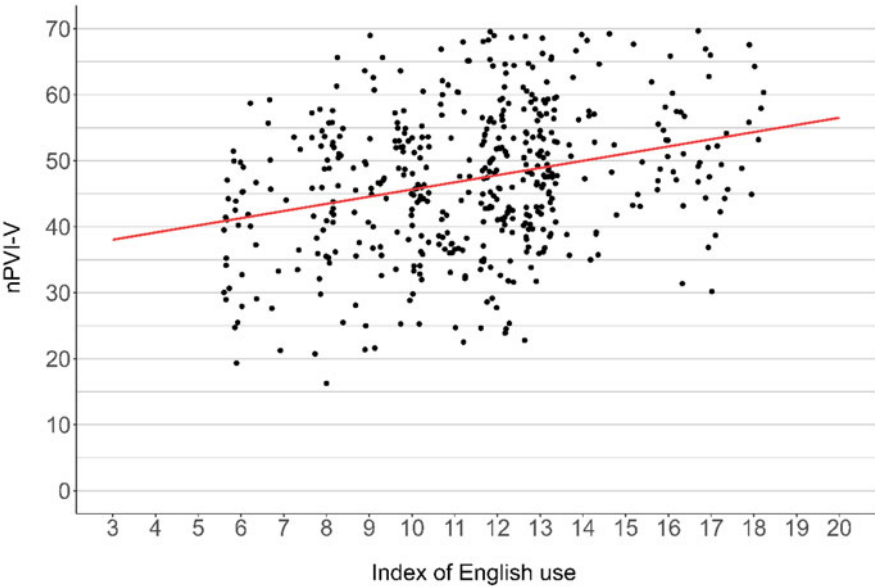
Figure 3. nPVI-V scores for speakers of Tongan English (N = 48) in informal conversational speech according to index of English use. Higher nPVI-V scores indicate tendency towards stress-timing, while lower nPVI-V scores indicate tendency towards syllable-timing.
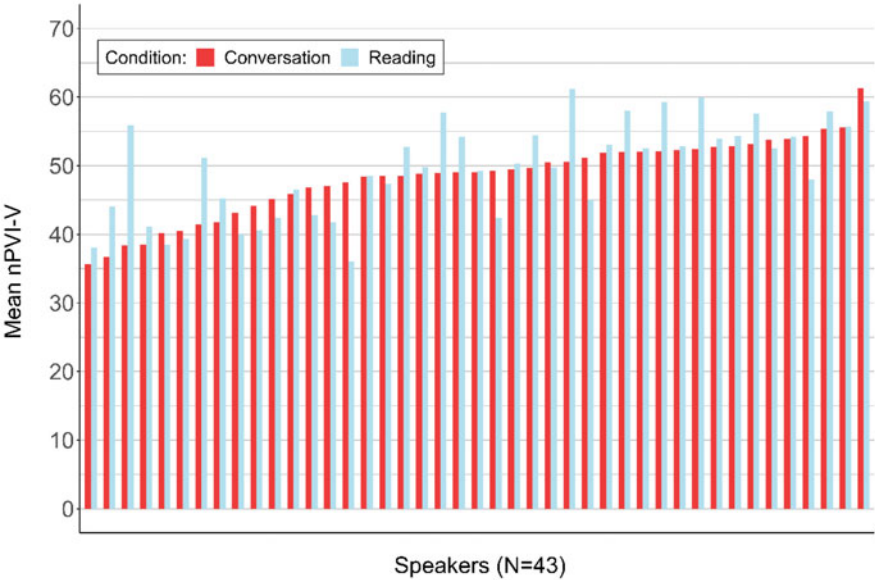

Figure 4. Mean nPVI-V scores for speakers of Tongan English (N = 43) in conversational and read speech conditions. Speakers are ordered from lowest nPVI-V in conversation to highest. Higher nPVI-V scores indicate tendency towards stress-timing, while lower nPVI-V scores indicate tendency towards syllable-timing.

for the majority of speakers (N = 27), while the remaining speakers (N = 16) exhibit higher nPVI-V scores in the conversational condition. In order to determine whether this observed intraspeaker variation according to speech condition was statistically significant, a linear regression model for nPVI-V was generated in which an interaction effect between *speaker* and *speech condition* was included. Although this returned as better-fit in comparison to the null model, a significant interaction effect was only found for one speaker (Sp30: $p < .05$), who exhibited greater stress-timing in read speech than conversation. Closer inspection does not point to any particular characteristic that could explain the significant result for this speaker.

These results for speech style differ somewhat from those of previous studies. On the whole, speakers tend to move towards syllable-timing in read speech in comparison to spontaneous or conversational speech (e.g. Gibbon & Gut 2001; Engstrand & Krull 2003; Szackay 2006; Arvaniti 2012). On the other hand, Leeman, Kolly & Dellwo (2014) also found great idiosyncrasy across speakers in read and spontaneous speech conditions, with some moving towards syllable-timing in read speech and others towards stress-timing.

One explanation is that the presence of a written orthographic text may have prompted some speakers to fully articulate vowels in unstressed syllables that would typically be reduced in spontaneous speech, affecting interval duration and thus nPVI-V scores, depending on how speakers were affected by the presence of a text.

The reading task could also be understood as a speech event in which speakers perform idiosyncratic narrative styles; some embodied a more theatrical reading style in which certain words were exaggerated through the lengthening of vowels, while others adopted a more monotonous style in reading aloud. A number of scholars have cited individual stylistic practice and performance as grounds for phonetic variation in reading tasks (e.g. Stuart-Smith *et al.* 2013; Gafter 2016; Wan 2022). In the current context, confidence in one's reading ability and familiarity with this practice may also contribute to variation in speech rhythm. Lengthening of vowels in narrative style may have increased variability in vocalic interval duration in speakers that are typically more syllable-timed in spontaneous speech, explaining the shift towards stress-timing in read speech for the majority of speakers.

3.5 *Acoustic analysis of vowel reduction*

Vowel reduction in unstressed syllables was analysed acoustically for eight speakers, four of whom exhibit the lowest nPVI-V scores in the corpus of conversational data and four of whom exhibit the highest nPVI-V scores, in order to determine the degree to which rhythmic differences could be attributed to this phenomenon. As outlined in section 1.1, syllable-timing is often attributed to the presence of fuller vowels in unstressed syllables leading to more equal syllable duration, while stress-timing is often attributed to vowel centralisation in unstressed syllables leading to greater variation in syllable duration. Central tendency was calculated for each speaker by measuring the average distance of F1 and F2 from the vowel centroid. Vowel centroids

were calculated using the following formula devised by Fabricius, Watt & Johnson (2009), a modified version of the Watt & Fabricius algorithm (2002) in which skewing of normalised values in the lower vowel space is avoided:[4]

$$S(Fi) = (Fi \text{ /i/} + Fi \text{ /ɑ/} + Fi \text{ /u'/}) / 3$$

To determine the vowel triangle parameters and vowel centroid for each speaker, eight tokens each of FLEECE and START were extracted for each speaker from conversational speech. A further ten tokens per speaker were extracted from unstressed syllables. F1 and F2 values were measured at three points (25%, 50%, 75%) in Praat and subsequently averaged. Values were normalised for the corpus following the Watt & Fabricius modified method, using the online software 'NORM' (Thomas & Kendall 2007). The resulting corpus contains F1 and F2 values for 80 reduced vowels (N = 8) and 64 tokens each of FLEECE and START from which the centroids and vowel space were calculated.

A linear regression model confirmed a significant difference between stress-timed and syllable-timed speakers for the distance of F1 from the centroid ($p = .0133$) but not for F2 ($p = .1528$). Essentially, this tells us that syllable-timed speakers are more likely to produce less centralised vowels in unstressed syllables in terms of vowel height (F1). This finding partially aligns with previous studies (Low *et al.* 2000; Jian 2004b; Kim & Lee 2005), which have found acoustic evidence for greater vowel centralisation in unstressed syllables of more stress-timed than syllable-timed speakers on the dimensions of both F1 and F2, while Fuchs (2016b) also found that F1 is an acoustic correlate of prominence but F2 is not.

4 Discussion

The current article has uncovered the social and stylistic constraints on variation in speech rhythm in TE, an emergent variety of English, by applying nPVI-V as a quantitative acoustic metric to a corpus of conversational and read speech data. As a variety, TE sits in the middle of the speech rhythm continuum in relation to other World Englishes that have been analysed using this metric, with a mean nPVI-V across the corpus (47.9, for conversational speech) closest to that of Singapore English (48.1, Tan & Low 2014: 209). Findings support the general tendency reported in World Englishes for emergent and L2 varieties to be more syllable-timed, while inner-circle L1 varieties are often more stress-timed (Mesthrie & Bhatt 2008), and the conceptualisation of rhythm as a continuum rather than a dichotomy.

Importantly, speakers of TE are by no means uniform and exhibit a great deal of variability in speech rhythm, with some resembling rhythmicity closer to L1 standard Englishes and others showing much more syllable-timed tendencies. This suggests that

---

[4] /u'/ is calculated based on measures of /i/ and /ɑ/, as tokens of /u/ (GOOSE lexical set) are commonly fronted. For further justification for and details of this calculation, see Watt & Fabricius (2002) and Fabricius *et al.* (2009).

classification of varieties of English should be done with caution, particularly in the case of emergent and L2 varieties, whereby speakers have varying degrees of proficiency and English use. This variation is discussed in the current section with regards to *occupation*, *education* (speaker social background) and *index of English use*, which returned as significant constraints on variation, considering the potential second-order indexicality of this variation in the local social context. Furthermore, variation in speech rhythm is discussed as a consequence of proficiency, language contact and L1 transfer, whereby degree of vowel reduction in unstressed syllables is an underlying linguistic mechanism affected by these factors.

Among the social factors that returned as significant constraints on rhythmic variation in TE were *occupation* and *education*, both of which are collinear and proxies for a locally salient distinction between local elite and other commoners. High degrees of stress-timing were exhibited by local elite commoners in the current study, which includes white-collar workers, students with white-collar aspirations and those with tertiary-level qualifications. During my fieldwork and interviews, local elite often reported using English more often on a daily basis, including in the social and private spheres. Drawing on Silversteen's (2003) concept of indexical order, stress-timed rhythm has the potential to be read as a second-order index of the local characteristics associated with eliteness in Tonga, such as high levels of education, wealth and leadership power (Moala 2009). This is not to suggest that stress-timing is consciously acquired by elite speakers as an index of these characteristics but comes to index such characteristics through its use by local elite. Stress-timing may thus take on social meaning in the local context as an index of education, wealth and power, as qualities that are associated with elite speakers who exhibit stress-timing.

One explanation for the tendency towards stress-timing in local elite speakers is rooted in the degree of exposure to English and level of proficiency. Elite TE speakers are using English in an increasing number of domains, including the home (Otsuka 2007), therefore the tendency towards stress-timing may be the result of higher proficiency in the language due to increased use and possibly earlier exposure. This argument is supported by findings in the current study for *index of English use*, whereby higher indexes predict a greater tendency towards stress-timing and lower indexes predict a greater tendency towards syllable-timing. Greater frequency of English use and exposure to English in formal education predicts proficiency and a greater likelihood of acquisition of stress-timed rhythm. Previous studies on speech rhythm have consistently revealed a relationship between the degree of instruction an individual has received in the target language and acquisition of the rhythmicity exhibited by L1 speakers of that variety, with rhythm moving closer to that of the target norm as the degree of instruction increases (e.g. Stockmal *et al.* 2005; Kim *et al.* 2006; White & Mattys 2007a, 2007b). Results from the current study reflect the same phenomenon, as individuals who have completed secondary education in a monolingual English-medium school exhibit more 'native-like' speech rhythm than those who have received a bilingual education or have not completed secondary school. These speakers show a lower degree of L1 influence as their level of proficiency and degree of English use increase. In addition, Tongan

royalty often exhibit features of metropolitan standards, in particular SSBE and NZE, which include a reduction of vowels in unstressed syllables and an impressionistic tendency towards stress-timing.[5] Members of the royal family commonly attend formal (secondary and tertiary) educational institutions abroad, often in the UK and New Zealand, where contact with speakers of an L1 metropolitan standard is high.

I argue that an exonormative model of standard speech is reinforced in Tonga when local elite speakers exhibit stress-timed tendencies. Following the argument that patterns in linguistic variants which align with a dominant social hierarchy also correspond to the evaluation of that variant as standard/non-standard (Labov 1972), we can infer that the use of stress-timing by these speakers reinforces this rhythmic tendency as standard, as illustrated in L1 metropolitan standards. This demonstrates ongoing support for the idea that inner-circle nations serve as norm providers to speakers in the outer and expanding circles (Kachru 1992). Observations of language use in the community and language ideologies in the full corpus of sociolinguistic interviews provide further support for the presence of an exonormative model in Tonga, whereby metropolitan standard varieties were often referred to as desirable while 'Tongan-sounding' English was often considered less attractive. Many informants expressed a standard language ideology, explicitly referring to the '*pālangi* accent' and L1 metropolitan standards such as GenAm, SSBE and NZE as desirable, which, as seen earlier in table 1, exhibit more stress-timed tendencies than emergent and L2 varieties. For example, in an interview with three young Tongans, the English spoken by those living in the US (i.e. GenAm) was reported as the most desirable variety of English, while TE was referred to as 'broken'. This sentiment of GenAm as appealing was more common amongst the younger generation, while middle-aged and older speakers often referred to SSBE, NZE and (less often) AusE as desirable. For example, one informant who expressed a preference for SSBE also reported a dislike for Māori NZE and the acquisition of this variety by Tongans in New Zealand, which according to previous studies exhibits more syllable-timed tendencies (e.g. Holmes 1997; Szakay 2006). Such discourse around ideal or desirable speech points to an exonormative model of speech in Tonga, reinforced and upheld by the local elite as a standard norm.

Lack of vowel reduction in unstressed syllables is often cited as a linguistic mechanism underlying syllable-timing, while vowel centralisation is linked to stress-timing (e.g. Lindblom 1963; Jian 2004b; Kim & Lee 2005; Mesthrie & Bhatt 2008). Empirical evidence for vowel centralisation was also found in the current study, albeit only in terms of vowel height (F1). In TE, more syllable-timed speakers often realise vowels in reduced syllables with a lower ʙᴀᴛʜ-like quality (e.g. *leader* as [ˈlidɐ]). It is important to note, however, that other measures of prominence also serve as cues to rhythmic grouping such as intensity (He 2012; Fuchs 2014a, 2016b), F0 (Cumming 2010; Fuchs

---

[5] This stress-timing is evident in interviews conducted with Tongan royalty that can be accessed online, as well as in the speech of a Tongan princess that I witnessed in person during my time in the field.

2014b) and sonority (Galves *et al.* 2002), and should be incorporated into durational metrics in the future.

The presence of fuller vowels in unstressed syllables is often linked to proficiency and considered an L1 transfer effect (e.g. Lee & Kim 2005; White & Mattys 2007a, 2007b; Sarmah *et al.* 2009), which is arguably the case in TE, given that vowel centralisation is not characteristic of secondary stress in Tongan (Garallek & White 2015), nor does the Tongan phoneme inventory contain central vowels (Feldman 1978). All informants in the current study were bilingual speakers with Tongan as an L1, thus the possibility of a transfer effect is likely. Melchers & Shaw (2003) note that speakers of L2 Englishes often do not acquire the complex vowel systems of L1 Englishes in which 20–24 different phonemes may be distinguished, including tense/lax distinctions, highlighting the difficulty in acquiring vowel patterns such as reduction and centralisation. It is possible that greater exposure to English at a younger age by TE speakers contributes to more successful acquisition of vowel reduction in unstressed syllables and therefore more stress-timed tendencies in speech. Age of acquisition, degree of instruction in the target language and level of engagement with English more generally appear to play an important role in predicting the degree to which a speaker will acquire vowel reduction in unstressed syllables and thus exhibit more stress-timed rhythmicity, while the social meaning of this tendency is apparent in stress-timing as an index of the characteristics associated with eliteness in Tonga.

5 Concluding remarks

The current article presented an acoustic analysis of speech rhythm in TE, determining the external and stylistic constraints on variation in rhythm by application of the metric nPVI-V. A tendency towards stress-timing was found in local elite speech, which may serve as a second-order index of the local characteristics associated with eliteness. Educated and white-collar local elite individuals exhibit greater tendencies towards stress-timing, while the speech of others is more syllable-timed. Furthermore, those with a higher index of English use are more likely to exhibit speech rhythm closer to the stress-timed end of the continuum. I have argued that exposure to English, particularly through formal education, and proficiency lead to a greater tendency towards stress-timing in TE. The linguistic mechanism underlying variation in rhythmicity can be explained, at least in part, by the degree of vowel reduction in unstressed syllables, as the most syllable-timed speakers of TE exhibit less centralisation of vowels with regards to F1 and thus more equal syllable durations, which can be attributed to an L1 transfer effect from Tongan.

The nPVI-V metric has been demonstrated in the current study as a valuable tool for the analysis of speech rhythm from a variationist perspective, revealing the social constraints on speech rhythm and illustrating gradient differences across speakers and between speech conditions. Importantly, findings suggest that we must proceed with caution when classifying emergent and L2 varieties of English on the basis of speech rhythm as stark discrepancies between speakers may exist that are conditioned by social factors

as well as idiosyncrasies in reading style. Future studies in World Englishes can profit from the application of nPVI-V, using a consistent method for the segmentation of vocalic and consonantal intervals in order to investigate inter- and intraspeaker variation in speech rhythm.

Author's address:

Department of English
University of Bern
Unitobler
Länggassstr. 49
3012 Bern
Switzerland
danielle.tod@unibe.ch

References

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Ainsworth, Helen. 1993. Rhythm in New Zealand English. Unpublished manuscript, Victoria University of Wellington.

Anderson, Victoria & Yuko Otsuka. 2006. The phonetics and phonology of 'definitive accent' in Tongan. *Oceanic Linguistics* 45, 21–42.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40, 351–73.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.

Bauer, Laurie. 1994. English in New Zealand. In Robert Burchfield (ed.), *The Cambridge history of the English language*, vol. 5: *English in Britain and overseas: Origins and development*, 382–429. Cambridge: Cambridge University Press.

Beinum, Florina. 1980. *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam: Academische Pers B.V.

Biewer, Carolin. 2020. Samoan English: An emerging variety in the South Pacific. *World Englishes* 40, 333–53.

Biggs, Bruce. 1971. The languages of Polynesia. *Current Trends in Linguistics* 8, 466–505.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341–5.

Bond, Dzintra, Dace Markus & Verna Stockmal. 2003. Prosodic and rhythmic patterns produced by native and non-native speakers of a quantity-sensitive language. In Maria-Josep Solé, Daniel Recasens & Joaquin Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 527–30. Adelaide: Causal Productions.

Bott, Elizabeth. 1981. Power and rank in the Kingdom of Tonga. *The Journal of the Polynesian Society* 90, 7–81.

Britain, David & Kazuko Matsumoto. 2012. Palauan English as a newly emerging postcolonial variety in the Pacific. *Language, Information, Text* 19, 137–67.

Carter, Phillip. 2005. Quantifying rhythmic differences between Spanish, English and Hispanic English. In Randall Scott Gess & Edward J. Rubin (eds.), *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages*, 63–75. Amsterdam and Philadelphia: John Benjamins.

Cumming, Ruth. 2010. The language-specific integration of pitch and duration. PhD dissertation, University of Cambridge.

Deterding, David. 2001. The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics* 29, 217–30.

Deterding, David. 2012. Issues in the acoustic measurement of rhythm. In José Romero-Trillo (ed.), *Pragmatics and prosody in English language teaching*, 9–24. Dordrecht: Springer.

Engstrand, Olle & Diana Krull. 2003. Rhythmic intentions or rhythmic consequences? Cross-language observations of casual speech. In Maria-Josep Solé, Daniel Recasens & Joaquin Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 2789–92. Adelaide: Causal Productions.

Fabricius, Anne, Dominic Watt & Daniel E. Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change* 21, 413–35.

Feldman, Harry. 1978. Some notes on Tongan phonology. *Oceanic Linguistics* 17, 133–9.

Fuchs, Robert. 2014a. Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English. In Nick Campbell, Dafydd Gibbon & Daniel Hirst (eds.), *Proceedings of the 7th International Conference on Speech Prosody*, 290–4. Dublin: ISCA.

Fuchs, Robert. 2014b. Towards a perceptual model of speech rhythm: Integrating the influence of f0 on perceived duration. In Li He, Helen Meng, Bin Ma, Eng Siong Chng & Lei Xie (eds.), *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 1949–53. Singapore: ISCA.

Fuchs, Robert. 2016a. *Speech rhythm in varieties of English: Evidence from educated Indian English and British English*. Singapore: Springer.

Fuchs, Robert. 2016b. The acoustic correlates of stress and accent in English content and function words. In Jon Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel & Nanette Veilleux (eds.), *Proceedings of the 8th International Conference on Speech Prosody*, 290–4. Boston, MA: ISCA.

Gafter, Roey. 2016. What's a stigmatized variant doing in the word list? Authenticity in reading styles and Hebrew pharyngeal. *Journal of Sociolinguistics* 20, 31–58.

Galves, Antonio, Jesus Garcia, Denise Duarte & Charlotte Galves. 2002. Sonority as a basis for rhythm class discrimination. In Bernard Bel & Isabel Marlien (eds.), *Proceedings of the 1st International Conference on Speech Prosody*, 323–6. Aix-en Provence: Laboratoire Parole et Langage.

Garellek, Marc & Marija Tabian. 2019. Tongan. *Journal of the International Phonetic Association* 50, 406–16.

Garellek, Marc & James White. 2012. Stress correlates and vowel targets in Tongan. *UCLA Working Papers in Phonetics* 110, 65–85.

Garellek, Marc & James White. 2015. Phonetics of Tongan stress. *Journal of the International Phonetics Association* 45, 13–34.

Gibbon, Dafydd & Ulrike Gut. 2001. Measuring speech rhythm. Paper presented at the 7th European Conference on Speech Communication and Technology, Aalborg, Denmark.

He, Lei. 2012. Syllabic intensity variations as quantification of speech rhythm: Evidence from both L1 and L2. In Qiuwu Ma, Hongwei Ding & Daniel Hirst (eds.), *Proceedings of the 6th International Conference on Speech Prosody*, 466–9. Shanghai: Tongji University Press.

Hess, Dominique. 2019. Saipanese English: History, structure and linguistic development. PhD dissertation, University of Bern.

Holmes, Janet. 1997. Maori and Pakeha English: Some New Zealand social dialect data. *Language in Society* 26, 65–101.

Jian, Hua-Li. 2004a. An improved pair-wise variability index for comparing the timing characteristics of speech. Paper presented at the 8th International Conference on Spoken Language Processing. Jeju Island, 4–8 October.

Jian, Hua-Li. 2004b. An acoustic study of speech rhythm in Taiwan English. Paper presented at the 8th International Conference on Spoken Language Processing, Jeju Island, 4–8 October.

Kachru, Braj. 1992. World Englishes: Approaches, issues and resources. *Language Teaching* 25, 1–14.

Kim, Jong-mi, Suzanne Flynn & Mira Oh. 2006. Non-native speech rhythm: A large-scale study of English pronunciation by Korean learners. 음성음운형태론연구 13, 219–50.

Kim, Jong-mi & Ok-hwa Lee. 2005. Reduced vowel quality accounts for Korean accent of English. *Studies on English Language and Literature* 31, 73–93.

Kisler, Thomas, Uwe Reichel & Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech and Language* 45, 326–47.

Knight, Rachael-Anne. 2011. Assessing the temporal reliability of rhythm metrics. *Journal of the International Phonetic Association* 41, 271–81.

Kuske, Eva. 2019. Guam English: Emergence, development and variation. PhD dissertation, University of Bern.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Lee, Ok-hwa & Jong-mi Kim. 2005. Syllable-timing interferes with Korean learners' speech of stress-timed English. *Speech Sciences* 12, 95–112.

Leemann, Adrian, Marie-José Kolly & Volker Dellwo. 2014. Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic Science International* 238, 59–67.

Lenth, R. 2018. *Emmeans: Estimated marginal means, aka least-squares means*. R package version 1.2.4. https://CRAN.R-project.org/web/packages/emmeans

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35, 1773–81.

Lloyd James, Arthur. 1940. *Speech signals in telephony*. London: Pitman and Sons.

Low, Ee Ling. 1998. Prosodic prominence in Singapore English. PhD dissertation, University of Cambridge.

Low, Ee Ling. 2006. A review of recent research on speech rhythm: Some insights for language acquisition, language disorders and language teaching. In Rebecca Hughes (ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice*, 99–125. Basingstoke: Palgrave Macmillan.

Low, Ee Ling & Esther Grabe. 1995. Prosodic patterns in Singapore English. In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences*, 636–9. Stockholm: KTH and Stockholm University.

Low, Ee Ling & Esther Grabe. 1999. A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech* 42, 39–56.

Low, Ee Ling, Esther Grabe & Francis Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43, 377–401.

Mesthrie, Rajend. 2008. Synopsis: The phonology of English in Africa and South and Southeast Asia. In Rajend Mesthrie (ed.), *Varieties of English: Africa, South and Southeast Asia*, 307–19. Berlin: de Gruyter.

Mesthrie, Rajend & Rakesh M. Bhatt. 2008. *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.

Melchers, Gunnel & Philip Shaw. 2003. *World Englishes*. London: Edward Arnold.

Moala, Kalafi. 2009. *In search of the Friendly Islands*. Kealakekua, HI: Pasifika Foundation Press.

Nishihara, Tetsuo & Jeroen Van de Wiejer. 2011. On syllable-timed rhythm and stress-timed rhythm in World Englishes: Revisited. 宮城教育大学紀要 56, 155–63.

Otsuka, Yuko. 2007. Making a case for Tongan as an endangered language. *The Contemporary Pacific* 19, 446–73.

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32, 693–703.

Pike, Kenneth. 1945. *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Ramus, Franck, Marina Nespor & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–92.

Riches, Christopher & Peter Stalker. 2016. *A guide to countries of the world*. Oxford: Oxford University Press.

Roach, Peter. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. In David Crystal (ed.), *Linguistic controversies*, 73–9. London: Edward Arnold.

Sarmah, Priyankoo, Divya Verma Gogoi & Caroline Wiltshire. 2009. Thai English: Rhythm and vowels. *English World-Wide* 30, 196–217.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23, 193–229.

Stockmal, Verna, Dace Markus & Dzintra Bond. 2005. Measures of native and non-native rhythm in a quantity language. *Language and Speech* 48, 55–63.

Stuart-Smith, Jane, Gwilym Pryce, Claire Timmins & Barrie Gunter. 2013. Television can also be a factor in language change: Evidence from an urban dialect. *Language* 89, 501–36.

Szackay, Anita. 2006. Rhythm and pitch as markers of ethnicity in New Zealand English. In Paul Warren & Catherine Watson (eds.), *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, 421–6. Canberra: Australasian Speech Science and Technology Association.

Tan, Rachel & Ee-Ling Low. 2014. Rhythmic patterning in Malaysian and Singapore English. *Language and Speech* 57, 196–214.

Taumoefolau, Melenaite. 1998. Problems in Tongan lexicography. PhD dissertation, University of Auckland.

Tent, Jan & France Mugler. 2008. Fiji English: Phonology. In Kate Burridge & Bernd Kortmann (eds.), *Varieties of English 3: The Pacific and Australasia*, 234–66. Berlin: Mouton de Gruyter.

Thomas, Erik & Phillip Carter. 2003. A cross-ethnic comparison of rhythm in the American South. Paper presented at the 4th UK Language Variation and Change conference.

Thomas, Erik & Phillip Carter. 2006. Prosodic rhythm and African American English. *English World-Wide* 27, 331–55.

Thomas, Erik & Tyler Kendall. 2007. *NORM: The vowel normalization and plotting suite* [software]. http://lingtools.uoregon.edu/norm/norm1.php

Tonga Statistics Department. 2019. *Tonga 2016 census of population and housing*, vol. 2: *Analytical report*. https://tongastats.gov.to/download/60/2016/3746/2016-census-report-volume-2.pdf

Torgensen, Eivind & Anita Szakay. 2011. A study of rhythm in London: Is syllable-timing a feature of multicultural London English? *University of Pennsylvania Working Papers in Linguistics* 17, 165–74.

Wan, Tsung-Lun Alan. 2022. Individual variation in performing reading aloud speech among deaf speakers. *Linguistics Vanguard* 8, 291–303.

Watt, Dominic & Anne Fabricius. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1-F2 plane. *Leeds Working Papers in Linguistics and Phonetics* 9, 159–73.

White, Laurence & Sven Mattys. 2007a. Rhythmic typology and variation in first and second languages. *Amsterdam Studies in the Theory and History of Linguistic Science Series* 4, 237–57.

White, Laurence & Sven Mattys. 2007b. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35, 501–22.